References and Links

This chapter compiles relevant references and provides links to projects where the MULTEXT-East morphosyntax has been developed and applied to Slovene texts.

Projects, in which the system has been developed or applied

MULTEXT-East - Multilingual corpora and text tools for Central and East European langauges

JOS - Linguistic Annotation of Slovene: Methods and Resources

Communication in Slovene

Janes - Resources, Tools and Methods for the Research of Nonstandard Internet Slovene

Development of Slovene in a Digital Environment

Training corpora containing manually revised MULTEXT-East tags

Krek, Simon; et al., 2019, Training corpus ssj500k 2.2, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1210.

Arhar Holdt, Špela; et al., 2022, Training corpus SUK 1.0, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1747.

Lenardič, Jakob; et al., 2022, CMC training corpus Janes-Tag 3.0, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1732.

References

Erjavec, Tomaž; Fišer, Darja; Krek, Simon in Ledinek, Nina. 2010. The JOS Linguistically Tagged Corpus of Slovene. V: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valeta, Malta, Maj. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/139_Paper.pdf [PDF]

Erjavec, Tomaž. 2012. MULTEXT-East: morphosyntactic resources for Central and Eastern European languages. Language Resources and Evaluation, 46(1): 131–142. DOI: 10.1007/s10579-011-9174-8 [PDF]

Erjavec, Tomaž. 2017. MULTEXT-East. V (Nancy Ide, James Pustejovsky, ur.): Handbook of linguistic annotation. pp. 441-462. Springer. DOI: 10.1007/978-94-024-0881-2_17

Krek, Simon; Erjavec, Tomaž; Dobrovoljc, Kaja; Gantar, Polona; Arhar Holdt, Špela; Čibej, Jaka in Brank, Janez. The ssj500k training corpus for Slovene language processing. V: Fišer, D. in Erjavec, T. Jezikovne tehnologije in digitalna humanistika: zbornik konference: 24.-25. september 2020, Ljubljana, Slovenija. Ljubljana: Inštitut za novejšo zgodovino, 2020. Str. 24–33.

http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Krek-et-al_The-ssj500k-Training-Corpus-for-Slovene-Language-Processing.pdf [PDF]

Revision #11
Created 2023-11-13 13:43:17 UTC by Tina Munda
Updated 2023-11-15 11:26:55 UTC by Tina Munda